Breast Skin-line Detection using Dynamic Programming

C. A. Silva, C. G. Lima and J. H. Correia

Abstract— In this paper, we present a novel method to extract the breast skin-line based on dynamic programming. Skin-line extraction is an important preprocessing step in CAD systems; however, it is a challenging problem due to the presence of noise, underexposed regions, which results in a low contrast area near the skin-air interface, and artifacts such as labels. Our proposal utilizes the stroma edge to constrain searching for the border. In order to cope with noise, we consider several candidate points for the border interface which are obtained by the Laplace operator applied in pre-defined directions in the mammogram. The breast contour is obtained from the candidate points using a dynamic programming algorithm. This utilizes a criterion of optimality to obtain the optimum contour by minimization of a cost function.

The method was evaluated using 82 mammograms whose contour were manually extracted by a radiologist from the mini-MIAS database. The Polyline Distance Measure was evaluated for each contour selected with the proposed method, obtaining a mean error of 2.05 pixels and a standard deviation of 0.80.

I. Introduction

According to the American Cancer Society, breast cancer is the second leading cause of cancer death in women, being exceeded only by lung cancer.

Although mammograms can miss some cancers, they remain a very effective tool for early breast cancer diagnosis. The sensitivity of mammography can be improved between $15\% - 30\%$ by an independent second reading of mammograms [1]. Some researchers [2], [3] have even envisaged that computerized analysis of mammograms can provide an emulated secondary opinion, improving consistency. This have driven research in techniques to automatically analyze and interpret mammograms. When developing new techniques to analyze and interpret breast images in CAD systems, an essential preprocessing step is the precise demarcation of the breast region in mammograms. Information on the breast border is used in several stages: for instance, in restricting the region where the algorithms are applied, for density correction of peripheral breast tissue [4] or in locating the nipple, which is used in registering the mammogram [5]. These utilizations of the breast contour require that it should properly model the soft-tissue/air interface and preserve the nipple. However, the presence of noise, regions underexposed near the top and bottom of the mammogram, low contrast in the breast border whose intensities are near the background and the presence of artifacts, such as labels, challenge the precise extraction of the breast skin-line, making it a difficult task. Most of these limitations are absent in digital mammography and the breast border is also easier to detect; however, screen-film mammography is still

The authors are with the Department of Electronics, University of Minho, Portugal. Corresponding author: CSilva@dei.uminho.pt

commonly used and may continue to be competitive since the digital mammography systems are much more expensive. Also digital mammography have not been found superior to screen-film mammography in all type of breast tissues [6]; therefore, improvement for CAD systems for screen-film mammographys continue to be an important research area [7].

Several approaches to the demarcation of the breast border in mammograms are proposed in the literature. Mendez [8] utilizes a gradient based method to delineate the breast contour by splitting the image in regions and scanning it in different directions according to its location. Lou [9] assumes that the intensity values from the breast region near the skin-air interface is a monotonic decreasing function; he exploits this observation to extract an initial boundary point that is refined until he obtains the final border point. These points are then linked to obtain the breast contour. Ferrari [10] obtained an initial contour by thresholding an image enhanced with the logarithm operator, whose noise was reduced with morphological opening. This was then refined by the adaptive active deformable model. Sun [11] used an adaptive thresholding to obtain an initial contour which was refined by a curve extrapolation method. This consisted in using the Euclidean distance as constraint that was propagated to the upper and lower breast regions to obtain the complete skin-line estimate. Thiruvenkadan [12] obtained an initial contour using a fuzzy segmentation which was refined by a region-based active contour segmentation method. Wu [7] applied a dynamically adaptive thresholding to the gray level range in local regions of the breast periphery to obtain the initial contour. This was improved by using gradient information from Sobel filtering to obtain the final boundary.

In this paper we present a new method for extracting the breast skin-line contour from mammograms based on dynamic programming. This consists in searching among several possible contours for the optimum contour according to a cost function. The algorithm pipeline comprehends three main stages: i) extraction of an initial contour region; ii) selection of contour candidate points and iii) selection of the best candidates such that a smooth path is obtained that minimizes a cost function.

II. Methodology

Several researchers [13], [11], [9] have identified important properties about mammograms which we utilize in designing our method for the extraction of the breast contour: *a)* the breast air interface is a region of low gradient and may be masked by noise or other artifacts such as labels; *b)* the

uncompressed fat near the breast-air interface has a gradient that increases as the fat tissue approaches the centre of the breast; c) the intensity values from the breast region near the skin-air interface is a monotonic decreasing function; and finally, d) the distance between the breast skin-line and the smoothed edge of the stroma is usually regular.

The almost regular distance between the stroma edge and the breast skin-line (d) permits to define a region where we have a high degree of confidence on the detection of the breast skin-line edge; however, property a) warns us that the edge may masked by noise and by the proximity of artifacts. Therefore, departing from the proposal by Sun [11], we enlarge the region where the edge may be found and consider several near by candidates. Based on properties b) and c), we apply the Laplace operator in the region near the stroma edge in pre-defined directions; the positive values of the Laplacian indicate the position of candidate edge points. The selection of the final contour is obtained by casting the problem into a dynamic programming framework which, based on a criterion of optimality, will obtain the optimal breast skin-line contour.

In the next section we develop on the main components of our method.

A. Components of the detection method

The main components of our method are the following:

1) Image preprocessing: According to property a), the effect of the noise will be specially adverse in the soft-tissue/air interface due to the low pixel intensity; therefore, we filter the image with a gaussian filter ($\sigma = 4$) and enhance the contrast near the border with the logarithm operator. The image is then linearly scaled to the range $0 - 255$.

2) Extraction of the initial contour region: The stroma edge area and other internal regions has a bright intensity, while the fatty periphery area has a low intensity near the background intensities. This bimodal characteristic of the mammogram indicates that we may obtain the initial contour region by thresholding the image using Otsu's threshold [14], T_{OTSU}. We used a threshold of $0.5T_{OTSU}$ to reduce the searching space for breast border candidate points.

Three landmark points are extracted from the segmented image, fig. 1. The landmark points A and B indicate the extremity of the edge top contour region. This is regularly sampled and the sampling points are interpolated using linear interpolation. The landmark point C split the initial top edge contour in two sub-regions. These sub-regions define how the image is analyzed: in sub-region I, the image is scanned horizontally line by line, while in region II, the scanning is performed vertically column by column.

3) Selection of contour candidate points: At each scanning iteration, we obtain a slice of the image (top plot, fig. 2). Starting at point B, we define an interval where we have a high confidence that the breast edge is localized (top plot – region delimited by the gray balls, fig. 2). Since the edge point may be masked by noise or by the proximity of a label, we consider several candidates based on the assumption that if the true edge is masked, then another point may be found
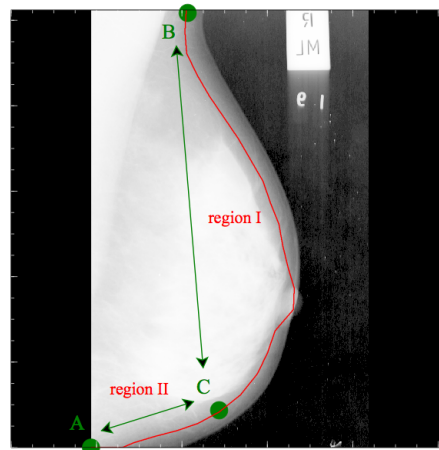

Fig. 1. Landmark control points (green) and initial top edge contour (red).

that is closer enough to the true one. These edge candidates are obtained by calculating each component of the Laplacian according to the scanning direction ($\partial^2 f/\partial x^2, \partial^2 f/\partial y^2$). The position of positive values of the Laplacian indicate edge candidates. Fig. 2 presents a region in the mammogram and a cut (middle plot) indicated by the yellow line, where the bottom plot presents the Laplacian on the region of interest.
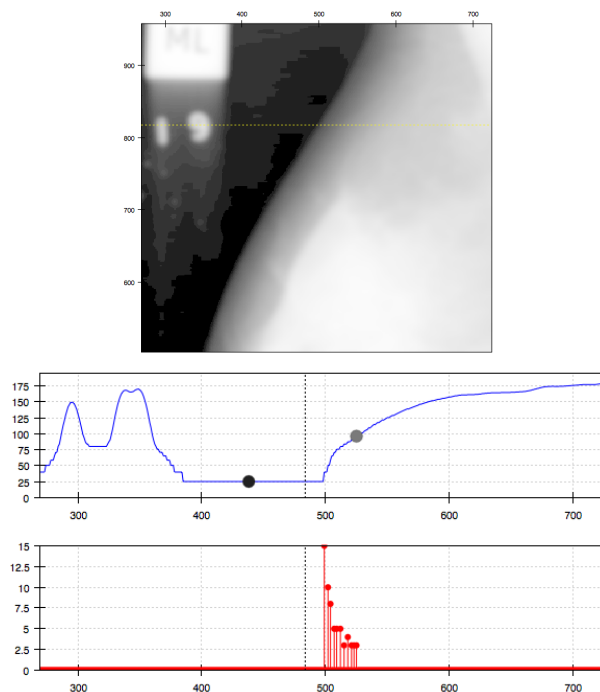

Fig. 2. Boundary candidate points.

4) Selection of the best candidates using dynamic programming: Given the edge candidate points, fig. 2, we may cast the problem of determining the skin-line border into a dynamic programming framework as follows: we are given a sequence of vectors, V_m ($m \in \{1, \ldots, M\}$), whose element $V_m(i)$ ($i \in \{1, \ldots, N\}$) is a pair of coordinates in the mammogram of a possible border point. We have

that N represents the number of candidates[1] and M the total number of points of the breast contour. The solution to our problem consists in finding a sequence of integers $j(m)$ ($m \in \{1, \ldots, M\}$) that represents the selections from the vectors V_m. This sequence is optimum according to a given criterion. This criterion of optimality is specified as the minimization of a cost function associated to every possible sequence.

This formulation of the problem may be described by a $N \times M$ matrix of nodes. The nodes in the mth column represent the N possible candidates from the neighborhood of the border contour. A selected sequence may be represented by a connected path along the columns.

A point in the border of the breast should have a low intensity value, so our optimal sequence $j(m)$ should minimize the total cost along the border. We define the total cost of the sequence terminating at candidate point j as

$$
\begin{aligned}
C(j(M), M) =\ & \alpha I[V_1(j(1))] + \sum_{m=2}^{M} \{ \alpha I[V_m(j(m))] \\
& + \beta d_E \left(V_{m-1}(j(m-1)), V_m(j(m)) \right) \\
& + \gamma d_{nei} \left(V_m(j(m)) \right) \},
\end{aligned}
$$

where I is the intensity value of a candidate point. We define d_E as the Euclidean distance between two successive candidates and d_{nei} as the distance between a candidate and the candidate at its right. The weights α, β and γ define the contribution of each term to the total cost function.

The dynamic programming algorithm may be written as follows: Suppose that we have determined the optimum path to every node in the column $(m-1)$. The total cost accumulated for these paths is $C(j, m-1)$, for $j = 1, \ldots, N$, respectively. By inspecting the expression of the accumulated cost, we have that the optimum path that ends at node (k, m) is the one coming from the node $(j^*, m-1)$, where j^* is given by

$$
\begin{aligned}
j^* =\ & \operatorname{argmin}_{j \in [1,N]} [C(j, m-1) \\
& + \beta d_E \left(V_{m-1}(j(m-1)), V_m(j(m)) \right)].
\end{aligned}
$$

The total accumulated cost for this path is

$$
\begin{aligned}
C(k, m) =\ & C(j^*, m-1) \\
& + \beta d_E \left(V_{m-1}(j(m-1)), V_m(k) \right) \\
& + \gamma d_{nei} \left(V_m(k) \right).
\end{aligned}
$$

The optimum paths up to $m-1$ may be extended to m by repeating the previous equation for each value of k. At position M, we select the path ending in the node for which $C(k, M)$ is minimum. The complete optimum path (the final breast contour) is recovered by backtracking, since the previous node from which the optimum path arrives is known.

---

[1]For the sake of the argument, we assume that N is the same for the entire sequence of vectors.

5) Post-processing: Having the optimum contour, we apply a median filter followed by a linear interpolator to close the path, since due to the process of scanning by lines and columns, we obtain sometimes small gaps in the other direction.

III. RESULTS AND DISCUSSION

A fair evaluation of our method would require comparing it with other published methods with a common data set. This could be accomplished by implementing and testing them with our method in a common data set; however, given the lack of detail found in some of these methods, it would be very difficult to correctly implement them. Also Padayachee [15] found variations between the radiologists' borders for clear breast edges and significant variations for unclear breast edges. For these reasons, we have adopted a publicly available data set whose ground-truth (GT) contours were previously used in other works ([10], [11]), in order to have a common base for a fair comparison. The GT set consists of 82 breast contours of the mini-MIAS data set [16] drawn by a radiologist.

In order to have a common error measure, we adopted the Polyline Distance Measure (PDM) [17] that was used in those previous works. The PDM is defined as the closest distance of each estimated boundary point to the GT boundary. It is a symmetric mean error measure.

Fig. 3 shows the comparison of the estimated skin-line boundary obtained with dynamic programming and with the adaptive active deformable model [10]. Table I compares the mean error and standard deviation using the dynamic programming, the adaptive active deformable model [10] and the dependency approach [11]. The weights of each term of the cost function used in the tests were $\alpha = 0.5$, $\beta = 6.5$ and $\gamma = 5.0$. Considering the table, we verify that the dynamic programming approach presents a mean error of 2.05 which corresponds to an improvement of 37.5 % over the dependency approach and 58.3 % over the deformable model; also important, is the variance that is much lower than the other two methods, indicating that the dispersion of the error is also lower. Fig. 4 presents the results for a mammogram.

The method was implemented using the scripting language Python and tested in a MacBook Pro 2.8 GHz with 4 GB of RAM. The execution time varied between 5 to 10s. This difference is due to the variable number of boundary candidate points. The dynamic programming algorithm is sometimes avoided because an exhaustive search of a large space of solutions can result in a very slow execution; however, by defining a restricted searching region and considering only certain candidates (those obtained by the Laplacian), we could reduce our searching space, obtaining a fast implementation.

IV. CONCLUSIONS

We have proposed a novel method for the demarcation of the breast border. Our approach utilizes physiological features in the mammogram to define a narrow region

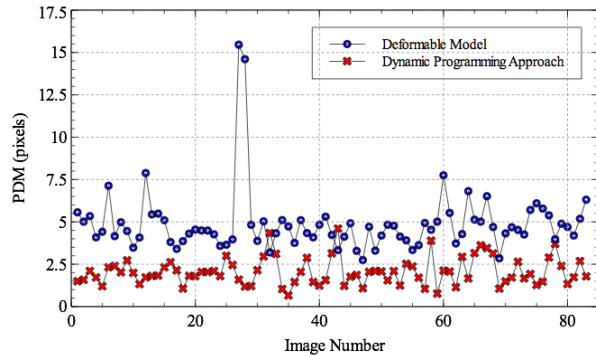|	μ (Pixels)	σ (Pixels)
Deformable model	4.92	1.91
Dependency approach	3.28	2.17
Dynamic programming approach	2.05	0.80

Fig. 3. PDM error curves.

near the breast skin-line border. By applying the Laplace operator we could define several possible candidates on the assumption that it permits to cope with noise and artifacts, then using dynamic programming, we obtained a optimum skin-line contour according to a cost function. The tests performed showed a good agreement between the contour obtained with our method and the ground-truth contours. Although we showed improved mean error and standard deviation in comparison to the two methods compared, further improvements must still account for a partial detection of the nipple in some of the mammograms. Experiments performed showed that this can be obtained by a proper selection of the weights of the cost function which should be obtained automatically.
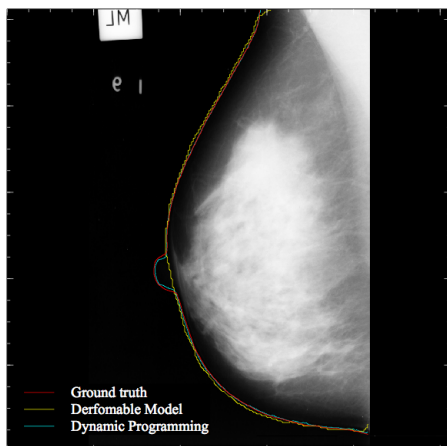

Fig. 4. Mammogram 'mdb003' from mini-MIAS.

REFERENCES

[1] E. Thurfjell, K. Lernevall, and A. Taube, "Benefit of independent double reading in a population-based mammography screening program," *Radiology*, vol. 191, pp. 241–244, 1994.
[2] R. Birdwell, K O'Shaughnessy D. Ikeda, and E. Sickles, "Mammographic characteristics of 155 missed cancers later detected with screening mammography and the potential utility of omputer-aided detection," *Radiology*, pp. 192–202, 2001.
[3] S. V. Destounis, P. DiNitto, W. Logan-Young, and E. Bonaccio, "Can computer-aided detection with double reading of screening mammograms help decrease the false-negative rate ? initial experience," *Radiology*, , no. 232, pp. 578–584, 2004.
[4] U. Bick, M. Giger, R. Schmidt, R. Nishikawa, and K. Doi, "Density correction of peripheral breast tissue on digital mammograms," *Radiographics*, vol. 16, no. 6, pp. 1403–1411, 1996.
[5] R. Chandrasekhar and Y. Attikiouzel, "A simple method for automatically locating the nipple mammograms," *IEEE Transactions on medical imaging*, vol. 16, no. 5, pp. 483–494, 1997.
[6] E. D. Pisano, C. Gatsonis, E. Hendrick, and M. Yaffe, "Diagnostic performance of digital versus fim mammography for breast-cncer screening," *N. Engl. J. Med.*, vol. 353, pp. 1773–1783, 2005.
[7] Y. T Wu, C. Zhou, H. P Chan, C. Paramagul, and L. M Hadjiiski, "Dynamic multiple thresholding breast boundary detection algorithm for mammograms," *Medical physics*, vol. 37, no. 1, pp. 391–400, 2010.
[8] A. Mendez, P. Tahoces, M. Lado, M. Souto, J. Correa, and J. Vidal, "Automatic detection of breast border and nipple in digital mammograms," *Computer methods and programs in biomedicine*, vol. 49, no. 3, pp. 253–262, 1996.
[9] S. Lou, H. Lim, K. Lin, and D. Hoogstrate, "Automatic breast region extraction from digital mammograms for pacs and telemammography applications," *Computerized Medical Imaging and Graphics*, vol. 24, no. 4, pp. 205–220, 2000.
[10] R. Ferrari, A. Frere, R. Rangayyan, J. Desautels, and R. Borges, "Identification of the breast boundary in mammograms using active contour models," *Medical and Biological Engineering and Computing*, vol. 42, no. 2, pp. 201–208, 2004.
[11] Y. Sun, J. Suri, J. Desautels, and R. Rangayyan, "A new approach for breast skin-line estimation in mammograms," *Pattern Analysis and Applications*, vol. 9, no. 1, pp. 34 – 47, 2006.
[12] S. Thiruvenkadam, M. Acharyya, N.V. Neeba, P. Jhunjhunwala, and S. Ranjan, "A region-based active contour method for extraction of breast skin-line in mammograms," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 2010, pp. 189–192.
[13] M. A Wirth and A. Stapinski, "Segmentation of the breast region in mammograms using snakes," in *Proceedings of the First Canadian Conference on Computer and Robot Vision*, 2004.
[14] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transaction on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979.
[15] J. Padayachee, M. J. Alport, and W. I. D. Rae, "Identification of the breast edge using areas enclosed by iso-intensity contours," *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 390–400, 2007.
[16] J. Suckling, J. Peter, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, and S. Kok, "The mammographic image analysis society digital mammogram database," in *2nd International Workshop on Digital Mammography*, 1994, pp. 375–378.
[17] J. S Suri, R. M Haralick, and F. H Sheehan, "Greedy algorithm for error correction in automatically produced boundaries from low contrast ventriculograms," *Pattern Analysis & Applications*, vol. 3, no. 1, pp. 39–60, 2000.